

NASH IMPLEMENTATION USING UNDOMINATED STRATEGIES

BY THOMAS R. PALFREY AND SANJAY SRIVASTAVA¹

We study the problem of implementing social choice correspondences using the concept of undominated Nash equilibrium, i.e. Nash equilibrium in which no one uses a weakly dominated strategy. We show that this mild refinement of Nash equilibrium has a dramatic impact on the set of implementable correspondences. Our main result is that if there are at least three agents in the society, then any correspondence which satisfies the usual no veto power condition is implementable unless some agents are completely indifferent over all possible outcomes. Many common welfare criteria, such as the Pareto correspondence, and several familiar voting rules, such as majority and plurality rules, satisfy our conditions.

This possibility result stands in sharp contrast to the more restrictive findings with implementation in either Nash equilibrium or subgame perfect equilibrium. We present several examples to illustrate the difference between undominated Nash implementation and implementation with alternative solution concepts.

KEYWORDS: Implementation, decentralization, social choice, equilibrium refinements.

1. INTRODUCTION

THE IMPLEMENTATION PROBLEM is to design a game such that a prespecified welfare criterion is guaranteed to be achieved by the game across a large domain of possible environments. As such, the implementation problem is fundamental to economics and its related social science disciplines. The study of this problem combines the essential ingredients of game theory, social choice theory, and the theory of incentives to rigorously analyze the role institutions play in the organization of economic, political, and social activity. Because the implementation problem is posed in terms of organizational design, its study has implications of an applied nature and is directly relevant to policy problems in regulation, negotiation and bargaining, and contract design. Because it lies at the juncture of several important theoretical subfields, it also provides insights into the logical foundations of basic theoretical constructs such as equilibrium concepts and social welfare functions.

It is therefore important to solve the implementation problem at a general level, that is, to characterize the boundaries of organizational design: exactly what can be implemented and according to which criteria of rational behavior.

The least controversial notion of rational behavior is given by dominant strategy equilibrium. In such an equilibrium, everyone uses a strategy which is a best response to *any* strategy profile played by others. Clearly, games which

¹ We thank the National Science Foundation for financial support. Palfrey also thanks the Exxon Educational Foundation and the Alfred P. Sloan Foundation for financial support while at the Center for Advanced Study in the Behavioral Sciences. We have benefited from participants of seminars at Caltech, Carnegie Mellon, Chicago, Michigan, Pennsylvania, Princeton, Stanford, SUNY-Buffalo, and USC. We also thank Dilip Abreu, Roger Guesnerie, Jean-Jacques Laffont, Paul Milgrom, John Moore, the referees, and especially Matthew Jackson, for helpful comments.

possess the property that there *exists* such an equilibrium across a broad domain of preferences have to be extremely special. This intuition is reflected in the negative results on dominant strategy implementation (Gibbard (1973), Satterthwaite (1975)); Dasgupta, Hammond, and Maskin (1979) provide a summary of results in this area).

These negative results led quite naturally to the widespread use of the much stronger form of rationality embodied in Nash equilibrium. Nash equilibrium requires that each agent use a strategy which is a best response to the specific strategies being used by the other agents, which is a form of “strategic rationality” not required by dominant strategy equilibrium. Comprehensive surveys of Nash implementation can be found in Maskin (1986), Postlewaite (1986), and Groves and Ledyard (1988).

Maskin (1977) provides a characterization of Nash implementable social choice correspondences (SCCs). He showed that a condition called monotonicity of an SCC was necessary and essentially sufficient for implementation. (See also, Repullo (1987), Saijo (1988), and Williams (1984).) Perhaps the most important implication of his characterization is that many reasonable welfare criteria fail to be implementable because of an indeterminacy or multiple equilibrium problem associated with Nash equilibrium. Thus, in contrast to dominant strategy equilibrium, the opposite problem arises with Nash equilibrium: there are too many equilibria. The key to obtaining positive results in Nash implementation theory thus involves mechanism design techniques which eliminate undesirable equilibria.

In a recent paper, Moore and Repullo (1988) have pursued an insight exploited by Moulin (1979), Crawford (1979), and Reichelstein (1987) that refinements of Nash equilibrium permit the implementation of some additional SCCs. They demonstrate that adding the restriction of sequential rationality partially alleviates the multiple equilibrium problem and thereby substantially expands the set of welfare criteria which can be implemented by carefully adding additional “veto” stages to the original game. (Their nearly complete characterization of this solution, called *Subgame Perfect Implementation*, has been extended by Abreu and Sen (1990).) They illustrate several applications of their result through a series of examples of incentive problems and identify welfare criteria which cannot be implemented via Nash equilibria, but which can be implemented via subgame perfect equilibria.

This paper also looks at refinements, but examines normal form refinements based on a simple dominance principle rather than exploiting sequential rationality. The results, perhaps surprisingly, are much stronger. Specifically, we make two assumptions about rational individual behavior which combine the features of dominant strategy implementation and Nash implementation. The first assumption, in the spirit of dominant strategy equilibrium, is completely “nonstrategic” and simply says that if, regardless of what actions others might be taking, one is never better off taking action *A* instead of action *B* and sometimes strictly better off taking action *B*, then action *A* is not taken. In the parlance of game theory, weakly dominated, or “inadmissible” strategies are not

used. The second assumption is the familiar best response (Cournot-Nash) criterion: given what actions others are taking, one chooses an action which does at least as well as any other action. We call this *Undominated Nash Implementation*. The restriction to Nash equilibria consisting only of admissible strategies is a very weak normal form refinement.

We then pose the implementation question relative to these assumptions about behavior. What SCC's are consistent with this type of behavior in the sense that there exists an institution under which this type of behavior will exactly reproduce the social choice correspondence? Our main result is that with three or more agents, all social choice correspondences satisfying the usual no veto power condition are implementable unless some agents are completely indifferent over all possible alternatives. This finding, besides providing an equilibrium notion for which the implementation problem with complete information is essentially solved, also can be viewed as the natural extension to general environments of our earlier paper in which we restricted attention to economic environments (Palfrey and Srivastava (1986a)).

Our results complement the findings of Moore and Repullo in several ways. First, what they emphasize is that the addition of extra stages in conjunction with additional rationality assumptions (sequential rationality) accomplishes more than the familiar method of adding extra strategies in a single stage game. What we show in this paper is that if weakly dominated strategies are eliminated, then the addition of extra stages is superfluous. Second, if there are more than two players, then any welfare criterion which satisfies no veto power and is implementable via subgame perfect equilibrium is also implementable in a single-stage game using undominated Nash equilibrium. Third, many welfare criteria which are implementable via undominated Nash equilibria are not implementable via subgame perfect equilibrium.

The next section contains the model and a number of important examples of welfare criteria which fail to be implementable in Nash equilibrium or subgame perfect equilibrium and explain the reasons for these failures. In Section 3, we identify a necessary condition for implementation via undominated Nash equilibrium and also prove that with three or more agents, it is sufficient for implementation of welfare criteria which satisfy the usual no veto power condition identified by Maskin (1977). For each of the examples of Section 2 we provide extremely simple games which successfully implement the desired outcomes using undominated Nash equilibrium. In Section 4, we mention a number of extensions of this result which involve both the relaxation of no veto power and the two person implementation problem. Some concluding comments are given in Section 5.

2. THE MODEL AND SOME EXAMPLES

2.A. *The Model*

There are I agents, indexed by $i = 1, 2, \dots, I$. We denote by \mathcal{A} the set of alternatives. A state (or an environment), denoted by s , specifies a profile of

preferences, one for each agent. S denotes the set of states, and the state is assumed to be common knowledge among the agents. In state s , the preference ordering of agent i is denoted by $R^i(s)$, which is a complete, reflexive, and transitive binary relation on A . $P^i(s)$ denotes the strict preference ordering derived from $R^i(s)$. Since states only distinguish preference profiles, we require that $s \neq s'$ implies $R^i(s) \neq R^i(s')$ for some i . With this convention, each state represents a unique preference profile, and there are no redundant states. Preferences over A are assumed to have maximal elements. For agent i in state s , we denote such an element by $b^i(s)$.

A *Social Choice Correspondence* (SCC) is a possibly multivalued mapping $F: S \rightarrow A$. For each state it specifies a set of alternatives.

A *mechanism* is a pair (M, g) , where $M = M^1 \times M^2 \times \cdots \times M^I$ and g is a function $g: M \rightarrow A$. M^i is the *message space* of agent i , and g is the *outcome function*. M^i serves as the strategy set of i at all $s \in S$. Let $M^{-i} = M^1 \times M^2 \times \cdots \times M^{i-1} \times M^{i+1} \times \cdots \times M^I$, with $m^{-i} \in M^{-i}$.

DEFINITION 1: $m^i \in M^i$ is a *best response* for i to $m^{-i} \in M^{-i}$ at state s if $g(m^{-i}, m^i) R^i(s) g(m^{-i}, \tilde{m}^i)$ for all $\tilde{m}^i \in M^i$.

DEFINITION 2: $m = (m^1, \dots, m^I) \in M$ is a *Nash equilibrium* at s if for all i , m^i is a best response to m^{-i} at state s .

If m is a Nash equilibrium of (M, g) at state s , then $g(m)$ is called a *Nash equilibrium outcome* at s . For a given mechanism, (M, g) , let $NE(s)$ denote the set of all Nash equilibrium outcomes at s .

DEFINITION 3: F is *implemented in Nash equilibrium* (by the mechanism (M, g)) if for all s , $F(s) = NE(s)$.

We turn next to Nash equilibria which do not involve the use of weakly dominated strategies. We call such equilibria undominated Nash equilibria (see van Damme (1987, p. 27)).

DEFINITION 4: A Nash equilibrium, m , is *weakly dominated* at s if there exists i and $\tilde{m}^i \in M^i$ such that

$$g(m^{-i}, \tilde{m}^i) R^i(s) g(m^{-i}, m^i) \quad \text{for all } m^{-i} \in M^{-i} \quad \text{and}$$

$$g(\tilde{m}^{-i}, \tilde{m}^i) P^i(s) g(\tilde{m}^{-i}, m^i) \quad \text{for some } \tilde{m}^{-i} \in M^{-i}.$$

This definition simply states that by playing the alternative strategy \tilde{m}^i , agent i is never worse off relative to playing m^i and he is strictly better off for some strategy combination of the other agents.

DEFINITION 5: $m \in M$ is an *undominated Nash equilibrium* (UNE) at s if m is a Nash equilibrium at s which is not weakly dominated at s .

Given a mechanism (M, g) , let $UNE(s) = \{g(m) | m \text{ is a UNE at } s\}$ denote the set of undominated Nash equilibrium outcomes at s .

DEFINITION 6: F is *implemented in undominated Nash equilibrium* (by the mechanism (M, g)) if for all s , $F(s) = UNE(s)$.

2.B. Examples

We turn next to three simple examples which illustrate that a number of important SCCs fail to be either Nash implementable or subgame perfect implementable.

For Nash implementation, Maskin (1977) identified monotonicity as a very simple and intuitive necessary condition for implementation:

DEFINITION 7: F is *monotonic* if for all x, s, s' satisfying $x \in F(s)$ and $x \notin F(s')$, there exist i, y such that $xR^i(s)y$ and $yP^i(s')x$.

Roughly speaking, this says that if x is in the social choice set at s , and the strict upper contour set relative to x does not expand for anyone at state s' , then x must also be in the social choice set at s' . To see why this is a necessary condition for Nash implementation is straightforward. Suppose that some game implements F , and $x \in F(s)$ is an equilibrium outcome at s . Since everything preferred to x at s' must also have been preferred to x at s for all agents of the game, the Nash equilibrium which generated x as an outcome at s must still be an equilibrium at s' . Since by assumption the game implements F , this implies that x lies in $F(s')$.

In other words monotonicity implies the following. Suppose one wishes to construct a mechanism to implement a social choice function for which x is the social choice at s , but not at s' . Monotonicity says that in order to do this there must exist an alternative y and some individual whose weak preference for x over y at s reverses to a strict preference for y over x at s' . Following the terminology introduced by Moore and Repullo (1988), we will refer to the pair (x, y) as a *test pair* for the states (s, s') .

If the domain of F includes a sufficiently large set of possible preference profiles and there are at least three possible outcomes, then monotonicity turns out to be extremely restrictive, the conclusion being that essentially no single valued social choice correspondence (i.e. a social choice function) is Nash implementable. With some fairly strong assumptions on the domain, some more positive results emerge with respect to multivalued correspondences. For example, the ("constrained") Walrasian correspondence is implementable in economic environments. However, in more general environments, even apparently reasonable correspondences are not Nash implementable. The following example shows this for the Pareto correspondence. (This example was pointed out to us by Faruk Gul.)

EXAMPLE 1—*Pareto Optimality*: There are 3 alternatives, $A = \{x, y, z\}$, 2 states, $S = \{s, s'\}$, and 2 agents. Preferences are given by:

$R^1(s)$	$R^1(s')$	$R^2(s) = R^2(s')$
x	xy	y
y	z	x
z		z

The Pareto correspondence evaluated at these two states is $F(s) = \{x, y\}$ and $F(s') = \{y\}$. Unfortunately, monotonicity requires that $x \in F(s')$, and F is therefore not implementable in Nash equilibrium. We remark that the Pareto correspondence is monotonic if S is restricted to strict orders on A or if $\{S, A\}$ correspond to neoclassical, pure exchange, economic environments. As a general rule, however, the Pareto correspondence is not monotonic.

EXAMPLE 2—*Plurality Rule* (from Abreu and Sen (1990)): There are 3 alternatives, $A = \{x, y, z\}$, 2 states, $S = \{s, s'\}$, and 3 agents. Preferences are given by:

$R^1(s) = R^1(s')$	$R^2(s) = R^2(s')$	$R^3(s)$	$R^3(s')$
x	y	z	y
y	z	y	z
z	x	x	x

Suppose the welfare criterion picks the alternative which is the first choice of the most number of agents, and otherwise uses an arbitrary tiebreaking procedure, say alphabetical order. Then we get $F(s) = x$ and $F(s') = y$. But monotonicity implies that we must have $x \in F(s')$.

EXAMPLE 3—*Majority Rule and Condorcet Winners*: There are 3 alternatives, $A = \{x, y, z\}$, 2 states, $S = \{s, s'\}$, and 5 agents. Preferences are given by:

$R^1(s) = R^2(s)$	$R^3(s)$	$R^4(s) = R^5(s)$	$R^1(s') = R^2(s')$	$R^3(s')$	$R^4(s') = R^5(s')$
x	y	z	x	yz	z
y	z	y	yz	x	y
z	x	x			x

Define an alternative w to be a *majority winner* against alternative w' if a majority of voters weakly prefer w to w' and more voters strictly prefer w to w' than w' to w ; w is called a *Condorcet winner* of A if it is a majority winner against every other alternative in A . This example is set up so that there is a unique Condorcet winner in each case with $F(s) = y$, $F(s') = z$. But monotonicity would require that $y \in F(s')$.

Many other examples can be constructed using more elaborate combinations of majority choice and scoring rules which yield nonmonotonic social choice

correspondences in the context of natural (and actually used) performance measures (see Palfrey and Srivastava (1986b) for further examples). These include virtually any runoff system for determining a winner in a multicandidate election when no clear majority winner exists, single and double elimination tournaments, the Borda count, and a variety of examples from private good economic environments such as those contained in Moore and Repullo (1988).

In fact the motivation for the extensive analysis of examples in Moore and Repullo (1988) is to show how much more powerful subgame perfect implementation is compared to (unrefined) Nash equilibrium. In the remainder of this section, we demonstrate that while refining Nash equilibrium by subgame perfection expands the set of implementable allocation rules, there still remains a large set of economically important welfare criteria which cannot be implemented. In particular, we show that none of the correspondences illustrated in Examples 1–3 are subgame perfect implementable. We then show in the next section that all three of them are implementable in undominated Nash equilibrium.

A precise necessary condition for subgame perfect implementation is given by Abreu and Sen (1990):

DEFINITION 8: F satisfies *Condition α* if there exists $B \subseteq A$ and $B \supseteq \text{range } F$ such that for all s and s' , if $x \in F(s)$ and $x \notin F(s')$, then there exists a finite sequence of allocations in B , $a_0 = x$, $a_1, a_2, \dots, a_l, a_{l+1}$ and a sequence of agents, $j(0), j(1), \dots, j(l)$ such that

- (i) $a_k R^{j(k)}(s) a_{k+1}$ for $k = 0, 1, 2, \dots, l-1$,
- (ii) $a_l R^{j(l)}(s) a_{l+1}$ and $a_{l+1} P^{j(l)}(s') a_l$,
- (iii) a_k is not $R^{j(k)}(s')$ maximal for $j(k)$ in B ,
- (iv) $j(l) \neq j(l-1)$ or $l = 0$, if a_{l+1} is $R^i(s')$ maximal in B for all $i \neq j(l)$.

This is a weaker condition than monotonicity since the test pair in (ii), (a_l, a_{l+1}) , need not have $a_l = x$, as would be required for monotonicity. However, a_l must be *indirectly* linked with x in a specific way, given by the other parts of Condition α .

This condition can be used to verify that the correspondences in the examples are not subgame perfect implementable. Consider Example 1. From part (ii) of condition α it must be the case that since $x \in F(s)$ but $x \notin F(s')$, there is an agent i and a pair of outcomes (b, c) such that $b R^i(s) c$ and $c P^i(s') b$. Since this is not the case for the SCC used in this example it follows immediately that the Pareto correspondence is not subgame perfect implementable. The Condorcet correspondence of Example 3 is not subgame perfect implementable for exactly the same reason. The SCC in Example 2 violates Condition α since for all sequences $j(0), \dots, j(l)$ and $x = a_0, a_1, a_2, \dots, a_{l+1}$, x is $R^{j(0)}(s')$ maximal for

$j(0)$. F is therefore not subgame perfect implementable. Additional examples which illustrate different ways in which scoring rules may fail to be subgame perfect implementable can be found in Palfrey and Srivastava (1986b).

3. NECESSARY AND SUFFICIENT CONDITIONS

This section contains our main results on UNE-implementable SCCs. We will show that an extremely weak condition, termed Property Q, is necessary for implementation. If there are at least three agents and no veto power holds, then it is also sufficient for implementation.

DEFINITION 9: F satisfies *Property Q* if for any s, s' , if $x \in F(s)$ and $x \notin F(s')$, then either

- (i) there exists i and $a, b \in A$ with $aP^i(s)b$ and $bR^i(s')a$ and there exist $c, d \in A$ with $cP^i(s')d$

or

- (ii) there exists i and $a, b \in A$ with $aR^i(s)b$ and $bP^i(s')a$.

This states that if $x \in F(s)$ and $x \notin F(s')$ then there must exist some agent with different preferences between the two states *and* that this agent cannot be completely indifferent between all alternatives at s' . This lack of complete indifference at s' need not be stated explicitly in (ii) since the statement $bP^i(s')a$ implies it.

It is straightforward to see that Property Q is necessary for undominated Nash implementation. If x is a UNE outcome at s and preferences at s' are either the same as at s or exhibit complete indifference, then x is also a UNE outcome at s' , so we must have $x \in F(s')$. Formally we state the following theorem.

THEOREM 1: *If F is UNE implementable, then F satisfies Property Q.*

Observe that Property Q is also a necessary condition for Nash implementation and for subgame perfect implementation.

Property Q is an extremely weak condition. To see this, observe that property Q is equivalent to Definition 9'.

DEFINITION 9': F satisfies *Property Q* if for any s, s' , if and only if for all i , $R^i(s) \neq R^i(s')$ implies for all $a, b \in A$, $aR^i(s')b$, then $x \in F(s) \Rightarrow x \in F(s')$.

This says that if x is an element of $F(s)$ and the only difference between preferences at s and s' is that at s' some agents are completely indifferent

between all alternatives, then x must also lie in $F(s')$. Consequently, it is not surprising that it is necessary for implementation. What is surprising is that with the (usual) additional requirements of three or more agents and no veto power, it is *sufficient* for UNE implementation. We now turn to this.

Recall that the definition of implementation of an SCC has two parts to it. The first part requires that every element of the SCC be an equilibrium outcome (“truthful implementation”), while the second part requires that every equilibrium outcome be an element of F . To implement an SCC, therefore, we have to ensure that at s , all elements of $F(s)$ are equilibrium outcomes and that nothing outside $F(s)$ ever arises in equilibrium. In the complete information environments we are analyzing, the first part is easy to accomplish; the difficulty arises in ensuring the second part. For example, suppose $x \in F(s)$ but $x \notin F(s')$. Then, we have to ensure that x is an equilibrium outcome at s but not at s' , which is impossible if the only preference changes are to complete indifference.

It is straightforward to see that (ii) of Condition α implies (ii) of Definition 9 and that Property Q is strictly weaker than Condition α . It is weaker in two ways. First, *all linkage between the “test pair” and x is eliminated*. Second, the test pair is not required to be a switch from a weak preference of a over b at s to a strict preference of b over a at s' . A switch of strict preference of a over b at s to a weak preference of b over a at s' is also acceptable.

Returning to the examples given earlier, we now demonstrate the basic intuition behind Property Q and illustrate the ideas that are used subsequently in the construction of mechanisms to fully implement via UNE.

We first show how the plurality rule of Example 2 can be implemented by a very simple extension of the message space. Here we exploit the fact that the test pair does not have to be linked back to x in the way required by Condition α . In the implementing mechanism, agents 1 and 2 have two possible messages, denoted by (δ, β) . Agent three has three possible messages, denoted (δ, β, γ) . The payoffs are:

		Agent 1				Agent 1	
		δ	β			δ	β
Agent 3	δ	x	z	δ	x	z	
	β	x	y	β	y	z	
	γ	x	z	γ	z	y	
		Agent 2 plays δ				Agent 2 plays β	

The unique undominated Nash equilibrium at s is (δ, δ, δ) . At s' , strategy δ is weakly dominated by strategy γ for agent 3, and there are two undominated Nash equilibria, (δ, β, β) and (β, β, γ) . Both equilibria yield y . The test pair in this example is (z, y) : agent 3 has a strict reversal of preferences between z and

y in the two states. Our construction exploits this reversal of preferences which permits the addition of a single message that dominates another message in state s' but not in s .

Next we show how the Pareto rule in Example 1 can be implemented via UNE. For this example, we exploit the second difference between Property Q and Condition α , namely that the reversal in preference in the test pair may go from strict preference to indifference. The following game, in which only agents 1 and 2 play, implements $F(s) = \{x, y\}$, $F(s') = \{y\}$:

		Agent 2	
		δ	β
Agent 1	δ	y	y
	β	z	x
	γ	y	z

In state s , the pair (δ, δ) and (β, β) are both undominated Nash equilibria. However, in state s' , (δ, δ) and (δ, β) are the only undominated Nash equilibria. The pair (β, β) is not a UNE at s' because δ weakly dominates β for agent 1. Using a similar construction, the Condorcet rule in Example 3 can be implemented via UNE even though it is not subgame perfect implementable (see Palfrey and Srivastava (1986b)).

In the above examples, we have shown how easy it is to construct implementing mechanisms using UNE when preferences over the test pair changes either from $aP^i(s)b$ to $bI^i(s')a$ (strict preference to indifference) or from $aP^i(s)b$ to $bP^i(s')a$ (strict preference reversal). The more difficult case to solve is a change in preference over the test pair from $aI^i(s)b$ to $bP^i(s')a$ (indifference to strict preference). The example below shows that for this case, in some instances we may have to construct a game with infinite strategy sets for the agents, even when there are only a finite set of states.

EXAMPLE 4— $A = \{x, y, z, w\}$, $S = \{s, s'\}$, $I = 2$. Both agents have identical preferences in each state, given by:

$R(s)$	$R(s')$
x	x
y	y
z, w	z
	w

Let $F(s) = \{x, y\}$, $F(s') = \{x\}$. Here, $y \in F(s)$ and $y \notin F(s')$. Part (ii) of Definition 9 is satisfied by agent 1 since $wR^1(s)z$ but $zP^1(s')w$. The following game implements F :

		Agent 2							
		α_0	α_1	α_2	α_3	α_4	α_5	α_6	\dots
Agent 1	α_0	x	z	z	z	z	z	z	\dots
	α_1	z	y	w	w	w	w	w	\dots
	α_2	z	y	z	w	w	w	w	\dots
	α_3	z	y	z	z	w	w	w	\dots
	α_4	z	y	z	z	z	w	w	\dots
	α_5	z	y	z	z	z	z	w	\dots
		\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots

There are an infinite number of Nash equilibria to this game at both s and s' : (α_0, α_0) yielding x as the outcome and (α_k, α_1) yielding y as the outcome for all $k \geq 1$. At s , none of these equilibria are weakly dominated, since both players are indifferent between z and w . At s' , however, for all $k \geq 1$, α_{k+1} weakly dominates α_k for player 1, since z is strictly preferred to w . The only undominated Nash equilibrium at s' is (α_0, α_0) , yielding x as the unique UNE outcome at s' . F is therefore UNE implementable. Note that if this game is truncated at any point, say at α_k for agent 1, then (α_k, α_1) becomes an undominated Nash equilibrium at s' .

The last example illustrates an interesting feature of our construction technique for eliminating undesirable equilibria in some very special situations. The construction technique for these cases uses an infinite chain of dominated strategies. This method is closely related to the “integer game” employed first by Maskin (1977) to eliminate extraneous Nash equilibria by constructing regions of the message space in which players are essentially engaged in a game to announce the highest integer. However, as we will see in the proof of Theorem 2, this type of construction is often unnecessary.

The previous examples illustrate how any change in preferences, except a change to complete indifference, can be used to eliminate undesired equilibria; in all of them, the problem was to eliminate y as an equilibrium outcome at s' . The mechanism constructed below to prove sufficiency is a general formulation of the intuition behind these examples. The proof also requires the following no veto power condition.

DEFINITION 10: F satisfies *no veto power* if for all s , whenever $aR^i(s)b$ for all $b \in A$ and for at least $I - 1$ agents, $a \in F(s)$.

No veto power is slightly stronger than unanimity. It says that if at some state all agents except possibly one share a most preferred alternative, then this alternative must be included in the social choice set at that state.

THEOREM 2: *If F satisfies Property Q and no veto power, and $I \geq 3$, then F is implementable in undominated Nash equilibrium.*

PROOF: A proof for the general case is given in the Appendix. Here, we outline a simple proof for the special case when all preferences are strict and there is a uniformly worst element outside the social choice set. That is, there exists $w \in A$ such that $w \notin F(s)$ for all s and for all i , for all s , $aP^i(s)w$ for all $a \neq w$. The intuition behind the construction is quite simple. We want to ensure that if $s' \neq s$ is the true state, and $x \notin F(s')$, then x is not a UNE outcome at s' . Since $s \neq s'$, Property Q implies there exist i, y_1, y_2 with $y_1P^i(s)y_2$ and $y_2P^i(s')y_1$. Call this agent $i(s, s')$ and the two allocations $y_1(s, s')$ and $y_2(s, s')$. We call i a *test agent* for the test pair $y_1(s, s'), y_2(s, s')$. The mechanism exploits this by including a set of (disequilibrium) joint messages at which i may switch the outcome from y_1 to y_2 . This is done as follows. First, $M^j = A \times S \times S \times \{-1, 0, 1, \dots, I\}$ is the message space of j . The typical equilibrium joint message at s will have $m^j = (x, s, t, 0)$ for all j with $x \in F(s)$, where either $t = s$ or $j = i(s, t)$.

The outcome function is constructed so that at s' , the situation facing $i(s, s')$ is essentially as follows:

		all $j \neq i(s, s')$	
		$(x, s, s, 0)$	$(x, s, s', -1)$
$i(s, s')$	$(x, s, s, 0)$	x	$y_1(s, s')$
	$(x, s, s', 0)$	x	$y_2(s, s')$

Then, it is not an equilibrium for all agents to report $(x, s, s, 0)$ when the state is s' ; in particular, $(x, s, s', 0)$ weakly dominates $(x, s, s, 0)$ for the test agent $i(s, s')$. This is how we exploit Property Q. The rest of the construction ensures that this type of dominance is preserved at s' for all other messages of the other agents. Finally, the integers in the message space ensure that the only possible equilibrium outcomes besides the typical ones satisfy no veto power. Formally, let

$M = M^1 \times M^2 \times \cdots \times M^I$. To define g , we divide M into a number of regions:

$$\begin{aligned}
 D_1 &= \{m \mid \exists s \in S, x \in F(s) \text{ such that for all } i, \text{ either } m^i = (x, s, s, 0) \\
 &\quad \text{or } i = i(s, s') \text{ and } m^i = (x, s, s', 0)\}, \\
 D_2 &= \{m \notin D_1 \mid \exists s \in S, x \in F(s) \text{ such that } m^j = (x, s, s, 0) \forall j \neq i\}, \\
 D_{3A}^i &= \{m \mid \exists s \in S, x \in F(s) \text{ with } m^j = (x, s, s', -1) \text{ for all } j \neq i; \\
 &\quad m^i = (x, s, s, 0) \text{ and } i = i(s, s')\}, \\
 D_{3B}^i &= \{m \mid \exists s \in S, x \in F(s) \text{ with } m^j = (x, s, s', -1) \text{ for all } j \neq i; \\
 &\quad m^i = (x, s, s', 0) \text{ and } i = i(s, s')\}, \\
 D_4 &= \{\text{all other } m\}.
 \end{aligned}$$

$$g(m) = \begin{cases} x & \text{if } m \in D_1, \\ w & \text{if } m \in D_2, \\ y_1(s, s') & \text{if } m \in D_{3A}^i, \\ y_2(s, s') & \text{if } m \in D_{3B}^i, \\ b^{i^*}(s^*) & \text{if } m \in D_4, \end{cases}$$

where $i^* = 1 + \text{mod}_I \{\sum_{j=1}^I \max\{0, m_4^j\}\}$ and $s^* = m_2^{i^*}$.

The power of this mechanism lies in regions D_{3A}^i and D_{3B}^i . There, all agents other than $i(s, s')$ report $(x, s, s', -1)$ while $i(s, s')$ reports $(x, s, s, 0)$ and the outcome is $y_1(s, s')$. If $i(s, s')$ switches his message to $(x, s, s', 0)$, we move to D_{3B}^i and the outcome changes to $y_2(s, s')$. By construction, $y_1(s, s')P^i(s)y_2(s, s')$ and $y_2(s, s')P^i(s')y_1(s, s')$. The outcome is unaffected by this switch by i elsewhere in the message space except possibly to move from D_2 to D_4 . Hence at s' , $m^i = (x, s, s, 0)$ is weakly dominated by $m^i = (x, s, s', 0)$. This means that all agents reporting $(x, s, s, 0)$ is not an undominated Nash equilibrium at s' . This weak domination does not arise at s because $y_1(s, s')P^i(s)y_2(s, s')$.

It is easy to verify that at s , for any $x \in F(s)$, $m^j = (x, s, s, 0)$ for all j is a Nash equilibrium which is not weakly dominated. Also, unless $m^j = (x, s, s, 0)$ for all j , there is a unilateral deviation which moves the aggregate message to D_4 and gives some j his best element. Consequently, if there is a Nash equilibrium other than $m^j = (x, s, s, 0)$ for all j , the outcome must be a best element for at least $I - 1$ agents, and no veto power implies the outcome is in $F(s)$. Finally, at $s' \neq s$, $m^j = (x, s, s, 0)$ for all j is not a UNE as argued above. Hence this mechanism implements F . Q.E.D.

This mechanism shares some features of the one used by Moore and Repullo to prove subgame perfect implementation, particularly the use of test pairs out of equilibrium. However, it differs in several important respects. First, and most

obviously, ours is a normal form game, and theirs is an extensive form game. Second, the normal form representation of their game is *not* equivalent to the one here. One major difference is that they require a (possibly very long) *sequence* of test pairs of allocations involving different individuals and different alternatives and must impose strong restrictions on the relationships between test pairs along such a sequence. While we concede that the use of predesignated test pairs at all seems a bit artificial, the use of a long predesignated sequence of such pairs is even more so.

The third, and perhaps most telling difference, is that our construction uses only a *single stage* of simultaneous play as opposed to many stages of simultaneous play. Consequently, we do not need to impose the additional restriction of sequential rationality in our equilibrium concept. In fact, the proofs for subgame perfect implementation rely heavily on the ability of each player to predict with perfect accuracy what the other players will do off the equilibrium path. Since there are multiple equilibria in some subgames of those mechanisms, this is a strong rationality requirement.

4. EXTENSIONS

In this section, we consider extensions of Theorem 2. Our first extension deals with relaxing the no veto power condition. While no veto power is a very weak restriction (for example, it is vacuously satisfied in pure exchange economies), it can sometimes impose undesirable restrictions on a welfare criterion. This is illustrated by the following example, which shows that no veto power can be inconsistent with Pareto optimality, and also that it is not necessary for UNE implementation.

EXAMPLE 5—*No Veto Power and Pareto Optimality*: There are 3 alternatives, $A = \{x, y, z\}$, 2 states, $S = \{s, s'\}$, and 3 agents. Preferences are given by:

$R^1(s)$	$R^1(s')$	$R^2(s) = R^2(s') = R^3(s) = R^3(s')$
x	x, y	x, y
y	z	z
z		

The Pareto optimal SCC for this example is given by $F(s) = \{x\}$, $F(s') = \{x, y\}$. No veto power requires that $y \in F(s)$, so that F does not satisfy the requirements of Theorem 2. However, F is implemented in undominated Nash equilibrium by the following trivial game:

	α	<div><div>x</div></div>
Agent 1	β	<div><div>y</div></div>

At s , α is the unique undominated Nash equilibrium yielding x as the

equilibrium outcome while at s' , both α and β are undominated Nash equilibria yielding x and y as UNE outcomes.

The following assumptions, which are weaker than no veto power, apply to situations similar to the one above. Let $B^i(s)$ and $W^i(s)$ denote the sets of best and worst elements for agent i at state s .

WEAK NVP: *For all s, s' , for all i , $W^i(s) \cap [\cap_{j \neq i} B^j(s')] \subset F(s')$.*

UNANIMITY: *For all s , $\cap_j B^j(s) \subset F(s)$.*

Weak NVP says that if a worst element for i at s is a best element at s' for all $j \neq i$, then this element lies in $F(s')$. Unanimity says that any unanimously best element at s must also lie in $F(s)$. Together, these assumption are weaker than no veto power.

THEOREM 3: *If F satisfies Property Q, Unanimity, Weak NVP, and $I \geq 3$, then F is implementable in undominated Nash equilibrium.*

PROOF: Same as that of Theorem 2 with the exception of Lemma 3 (in the Appendix). In Lemma 3, weak NVP yields that if there is an equilibrium at s in D_3^i for some i , the outcome lies in $F(s)$. Unanimity yields that if there is an equilibrium at s in $D_{4,A}^i$ to D_{10} , then the outcome lies in $F(s)$. *Q.E.D.*

Theorem 3 implies that with strict preferences, the Pareto correspondence is implementable, as is any selection from the Pareto correspondence which preserves weak NVP. Further, several voting rules including majority and plurality rules satisfy the hypotheses of Theorem 3.

We turn next to the case of two agents. In this case, we show that the additional domain assumption that there exists a unanimously undesirable alternative suffices for implementation. That is, there is an alternative which is ranked strictly last on *all preference orders in the domain*. Such an assumption has been used by McKelvey (1989) to obtain strategy space reductions for Nash implementation and was also used to illustrate the proof of Theorem 2 in this paper. A slightly weaker assumption is used by Moore and Repullo (1988) for subgame perfect implementation in the 2-agent case. We also show that with this assumption, no veto power can be replaced with unanimity, and the requirement of three or more agents can be dispensed with.

ASSUMPTION H: *There exists $w \in A$ such that for all i , for all s , $aP^i(s)$ for all $a \in A \setminus \{w\}$, and $w \notin F(s)$ for all s .*

THEOREM 4: *If F satisfies Property Q, Unanimity, H, and $I \geq 2$, then F is implementable in undominated Nash equilibrium.*

PROOF: If $I > 2$, then Theorem 3 applies. If $I = 2$, we modify the mechanism used in the proof of Theorem 2. Note first that with $I = 2$, D_3^i and D_3^j may have a nonempty intersection. We let $g(m) = w$ if $m \in D_3^i$ for any i . Second, since $x \in F(s)$ is never a worst element for any i and s , we let the outcome in D_{9A}^i and D_{9B}^j be determined as in D_{10} . It can now be checked that Lemmas 1, 2, 3, and 5 (in the Appendix) still hold. In Lemma 4, $m \in D_3^i$ is never an equilibrium outcome at any s since either agent can unilaterally move m to $D_1 \cup D_2$ and get $xP^i(s)w$. Finally, unanimity ensures that if m is an equilibrium at s in D_{4A}^i to D_{10} , $g(m) \in F(s)$. Q.E.D.

5. DISCUSSION AND CONCLUSIONS

Our results show that in general collective choice problems, virtually all welfare criteria are implementable in undominated Nash equilibrium. In this section we comment on some features of our analysis.

Undominated Nash equilibrium is a refinement of the Nash equilibrium concept which is weaker than Selten's (1975) notion of a trembling-hand perfect equilibrium (see van Damme (1987, p. 28) for a discussion of this point). One advantage of UNE over perfection is that it is easy to define for infinite games, while extensions of perfect equilibrium to infinite games can lead to weakly dominated strategies being played in equilibrium. In our setting, it is natural to work with infinite games. If the set of alternatives, A , is infinite (as would be the case, for example, if A was the set of feasible allocations in a pure exchange economy), then the set of preferences over A is also infinite, and any game which involves elicitation of preferences will naturally involve an infinite strategy space. In this sense, UNE seems a more natural concept of equilibrium for our setting than is perfection. It should also be noted that in the characterization of perfect equilibrium in terms of lexicographic Nash equilibrium (see Brandenburger and Dekel (1986), Blume (1986)), perfection requires strong assumptions on "higher order" beliefs while the relaxation of these assumptions yields undominated Nash equilibrium.

Our next comment concerns the size of the mechanism we construct in our sufficiency proof. The mechanism for the general case may appear to be unintuitive and complicated. These complications are due to the fact that we constructed a single mechanism to implement any SCC satisfying our conditions. Consequently, we were not able to use information which might be specific to a given problem. For example, the mechanism constructed in Palfrey and Srivastava (1986a) for pure exchange economies is much simpler than the one constructed here, but exploits several important features of the pure exchange model. The mechanism for strict preference domains given in Section 3 is also quite simple. Further, as the examples in Section 3 show, in many cases implementing mechanisms may be nearly as simple as "direct" mechanisms.

However, the mechanisms used for the general constructive proofs in virtually all of implementation theory tend to be cumbersome and artificial. In our case,

while eliminating weakly dominated strategies turns out to be a very powerful tool, the way in which this power is exploited raises some deep issues.

First, there is the issue of whether the rationality requirement that players do not use dominated strategies is a weak assumption or a strong assumption. Is a single round of elimination of dominated strategies a reasonable refinement criterion? If so, then it is important to characterize precisely what can be implemented and to give an indication of how to design mechanisms to do so. Our opinion on this is that *compared with the strength of the mutual best response criterion of Nash equilibrium and compared with any other refinement of Nash equilibrium*, eliminating one round of dominated strategies is a very weak assumption. This view is supported by recent work by game theorists on the rational foundations of noncooperative play (see, for example, Dekel and Fudenberg (1987)). It is, by most criteria, one of the weakest possible refinements of Nash equilibrium (van Damme (1987, p. 314, Figure 1)).

Second is the question of whether it is an appropriate criterion to apply if there are an infinite number of strategies. Here, we take the view that the kind of problems that may arise with respect to infinite strategy spaces are problems of a more fundamental sort that also challenge the use of the Nash equilibrium concept and most other refinements. For one thing, the rationality arguments cited above do not depend on there being a finite number of strategies. Potentially more problematic is that the undominated best response correspondence can be empty in infinite games. Of course, this can also happen with the best response correspondence, which can be empty when there are an infinite number of strategies, but for different reasons. Examples of this in economic applications abound, including spatial location games in which the “best response” is to move as close to one’s opponent as possible, or games of timing where one would like to move immediately before (or after) one’s opponent, or in bidding and pricing games, where one would like to bid slightly higher (or lower) than one’s opponent. But the nonexistence of best responses *off the equilibrium path* does not necessarily negate the value of applying the criterion of a mutual best response (Nash) equilibrium when one exists. Similarly, the nonexistence of undominated best responses off the equilibrium path need not negate the value of applying the criterion of undominated mutual best response (Nash) equilibrium when one exists.

Third is the question of whether it is appropriate to exploit an infinite chain of dominated strategies for purposes of mechanism design. In particular, our general proof (though not the proof with strict preferences under Assumption H) has an infinite chain of dominated strategies. This is a more difficult issue to answer, and we are not sure there is a good answer for the time being. Our tentative answer is that we are uncomfortable with this, in much the same way we are uncomfortable with the use of integer games and other artificial features found almost universally in the constructive proofs of implementation theorems. We must bear in mind that these are primarily existence theorems (including the ones in this paper), and that the first step towards proving an existence

theorem is finding a correct proof. Once existence of such mechanisms has been established, we can go on to study properties of these mechanisms. Under exactly what circumstances are infinite chains of dominated strategies required? Do “real world” mechanisms have features which are similar to features of these apparently abstract mechanisms? These and related questions are important subjects of future research, but are not the main focus of the present paper.

One interpretation of our results is that they provide a general possibility theorem and show that noncooperative behavior is consistent with an extremely large class of welfare criteria. Since part of the motivation for our analysis was that previously studied equilibrium notions were unable to implement several important welfare criteria, this is a strong positive result insofar as UNE allows us to implement most SCC's. The constructive proof provides one method for applying these general results to specific problems of mechanism design. One obvious direction to take this research is to investigate the existence of simple “realistic” mechanisms in specific applications.

Finally, we note that we have limited the analysis to situations of complete information. Jackson (1991), Palfrey and Srivastava (1987, 1989a), and Postlewaite and Schmeidler (1986), and others, provide a characterization of Bayesian implementable SCC's in settings of incomplete information, while a general characterization of undominated Bayesian implementable allocations is given in Palfrey and Srivastava (1989b).

*Division of Humanities and Social Sciences, California Institute of Technology,
Pasadena, CA 91125, U.S.A.*

and

*Graduate School of Industrial Administration, Carnegie Mellon University,
Pittsburgh, PA 15213, U.S.A.*

Manuscript received December, 1986; final revision received February, 1990.

APPENDIX

We prove the theorem first for the special case in which in each state, each agent has a worst element. At the end, we show how the proof changes when this is relaxed. For each i , let $w^i(s) \in W^i(s) = \{x | zR^i(s)x \text{ for all } z \in A\}$.

For any s and s' , if (i) of Definition 9 holds, define $i_1(s, s')$, $y_1(s, s')$, and $y_2(s, s')$ to be an agent and a pair of alternatives such that $y_1 P^i(s) y_2$ and $y_2 R^i(s') y_1$. If (ii) but not (i) of Definition 9 holds, define $i_2(s, s')$, $z_1(s, s')$, and $z_2(s, s')$ to be an agent and a pair of alternatives such that $z_1 R^i(s) z_2$ and $z_2 \bar{P}^i(s') z_1$. Let

$$M^i = M_1^i \times M_2^i \times M_3^i \times M_4^i \times M_5^i \quad \text{and} \quad M = M^1 \times \cdots \times M^I$$

where

$$\begin{aligned} M_1^i &= \{(x, s) | x \in F(s), s \in S\}, & M_2^i &= S, \\ M_3^i &= \{-4, -3, -2, -1, 0, 1, 2, \dots\}, & M_4^i &= \{0, 1, 2, \dots\}, & M_5^i &= S. \end{aligned}$$

We partition M as follows.

$$\begin{aligned}
 D_1 &= \{m | m^j = (x, s, s, 0, 0, s) \forall j\}, \\
 D_2 &= \{m | \forall j, m^j = (x, s, s, 0, 0, s), \text{ or } m^j = (x, s, s', -1, k^j, s^j) \text{ and } j = i_1(s, s'), \\
 &\quad \text{or } m^j = (x, s, s', -3, k^j, s^j) \text{ and } j = i_2(s, s'); m \notin D_1\}, \\
 D_3^i &= \{m | \forall j \neq i, m^j = (x, s, s, 0, 0, s), \text{ or } m^j = (x, s, s', -1, k^j, s^j) \text{ and } j = i_1(s, s'), \\
 &\quad \text{or } m^j = (x, s, s', -3, k^j, s^j) \text{ and } j = i_2(s, s'); m \notin D_1 \cup D_2\}, \\
 D_{4A}^i &= \{m | \forall j \neq i, m^j = (x, s, s'', -2, 0, s); i = i_1(s, s''); m^i = (x, s, s, 0, 0, s), \\
 &\quad \text{or } m^i = (x, s, s', -3, k^i, s^i) \text{ and } i = i_2(s, s')\}, \\
 D_{5A}^i &= \left\{m | \forall j \neq i, m^j = (x, s, s'', -2, k^j, s); \max_{j \neq i} k^j > 0; i = i_1(s, s''); \right. \\
 &\quad \left. m^i = (x, s, s, 0, 0, s), \text{ or } m^i = (x, s, s', -3, k^i, s^i) \text{ and } i = i_2(s, s')\right\}, \\
 D_{6A}^i &= \{m | \forall j \neq i, m^j = (x, s, s', -2, 0, s); m^i = (x, s, s', -1, k^i, s^i); i = i_1(s, s')\}, \\
 D_{7A}^i &= \{m | \forall j \neq i, m^j = (x, s, s'', -2, k^j, s); i = i_1(s, s''); m^i = (x, s, s', -1, k^i, s^i) \\
 &\quad \text{and } i = i_1(s, s'); m \notin D_{6A}^i\}, \\
 D_{8A}^i &= \{m | \forall j \neq i, m^j = (x, s, s'', -2, k^j, s); i \neq i_1(s, s''); m^i = (x, s, s, 0, 0, s), \\
 &\quad \text{or } m^i = (x, s, s', -1, k^i, s^i) \text{ and } i = i_1(s, s'), \\
 &\quad \text{or } m^i = (x, s, s', -3, k^i, s^i) \text{ and } i = i_2(s, s')\}.
 \end{aligned}$$

Regions D_{4B}^i to D_{8B}^i are defined analogously to regions D_{4A}^i to D_{8A}^i except that -1 is replaced by -3 , -2 by -4 , -3 by -1 , i_1 by i_2 , and i_2 by i_1 .

$$\begin{aligned}
 D_{9A}^i &= \{m | \forall j \neq i, m^j = (x, s, s, i, 0, s); m^i = (x, s, s, 0, 0, s), \text{ or } m^i = (x, s, s', -1, k^i, s^i) \\
 &\quad \text{and } i = i_1(s, s') \text{ or } m^i = (x, s, s', -3, k^i, s^i) \text{ and } i = i_2(s, s')\}, \\
 D_{9B}^i &= \{m | \forall j \neq i, m^j = (x, s, s, i, 0, s); m \notin D_{9A}^i\}, \\
 D_{10} &= \left\{m | m \notin D_1 \cup D_2 \cup \left[\bigcup_i D_3^i \cup \cdots \bigcup_i D_{9B}^i \right] \right\}.
 \end{aligned}$$

For $m \in D_{10}$, let

$$I^* = \{i | \max\{0, m_3^i\} \geq \max\{0, m_3^j\} \text{ for all } j\}$$

and $i^* = \min_{i \in I^*} \{i\}$. For $i_2(s, s'')$, let

$$z^*(s^i, s, s'') = \begin{cases} z_1(s, s'') & \text{if } z_1(s, s'') R^i(s^i) z_2(s, s''), \\ z_2(s, s'') & \text{otherwise,} \end{cases}$$

$$g(m) = \begin{cases} x & \text{if } m \in D_1 \cup D_2, \\ w^i(s) & \text{if } m \in D_3^i \cup D_{9B}^i, \\ y_1(s, s'') & \text{if } m \in D_{4A}^i, \\ w^i(s'') & \text{if } m \in D_{5A}^i \cup D_{8B}^i, \\ y_2(s, s') & \text{if } m \in D_{6A}^i, \\ b^i(s^i) & \text{if } m \in D_{7A}^i \text{ and } k^i > \max_{j \neq i} k^j, \\ w^i(s'') & \text{if } m \in D_{7A}^i \text{ and } k^i \leq \max_{j \neq i} k^j, \\ b^i(s^i) & \text{if } m \in D_{8A}^i \text{ and } m_3^i = -1 \text{ and } k^i > \max_{j \neq i} k^j, \\ w^i(s'') & \text{if } m \in D_{8A}^i \text{ and } m_3^i \neq -1 \text{ or } k^i \leq \max_{j \neq i} k^j, \\ z_1(s, s'') & \text{if } m \in D_{4B}^i \cup D_{5B}^i, \\ z_2(s, s') & \text{if } m \in D_{6B}^i, \\ z^*(s^i, s, s'') & \text{if } m \in D_{7B}^i \text{ and } k^i > \max_{j \neq i} k^j, \\ z_1(s, s'') & \text{if } m \in D_{7B}^i \text{ and } k^i \leq \max_{j \neq i} k^j, \\ b^i(s) & \text{if } m \in D_{9A}^i, \\ b^i(m_5^*) & \text{if } m \in D_{10}. \end{cases}$$

LEMMA 1: $x \in F(s)$ implies $m^j = (x, s, s, 0, 0, s)$ for all j is a Nash equilibrium at s .

PROOF: Any unilateral deviation by i moves the joint message from D_1 to either D_2 or D_3^i . In either case, i is no better off. Q.E.D.

LEMMA 2: $x \in F(s)$ implies $m^j = (x, s, s, 0, 0, s)$ for all j is not weakly dominated at s .

PROOF: Consider any i . There are three types of deviations i can make:

- (i) $\bar{m}^i = (x, s, s', -1, k^i, s^i)$ where $i = i_1(s, s')$ for some s' ,
- (ii) $\bar{m}^i = (x, s, s', -3, k^i, s^i)$ where $i = i_2(s, s')$ for some s' ,
- (iii) any other deviation.

Case (i): Consider $\bar{m}^i = (x, s, s', -2, 0, s) \forall j \neq i$. Then, $(\bar{m}^{-i}, m^i) \in D_{4A}^i$ while $(\bar{m}^{-i}, \bar{m}^i) \in D_{6A}^i$, so $g(\bar{m}^{-i}, m^i) = y_1(s, s')$ and $g(\bar{m}^{-i}, \bar{m}^i) = y_2(s, s')$. Since $y_1 P^i(s) y_2$, i is strictly worse off using \bar{m}^i .

Case (ii): First note that if $(x, s, s', -3, k^i, s^i)$ weakly dominates m^i at s , then so does $(x, s, s', -3, k^i, s)$, because the last element of i 's message is only used in D_{7B}^i to determine z^* and in D_{10} to determine i 's best element when $i = i^*$. Therefore, let $\bar{m}^i = (x, s, s', -3, k^i, s)$. For any \bar{m}^{-i} , consider $g(\bar{m}^{-i}, m^i)$ and $g(\bar{m}^{-i}, \bar{m}^i)$, and observe that the outcome is only affected when $(\bar{m}^{-i}, m^i) \in D_{4B}^i \cup D_{5B}^i$. If $\bar{m}^i = (x, s, s', -4, 0, s) \forall j \neq i$ and $i = i_2(s, s'')$ and $s'' \neq s'$, the outcome can change from $z_1(s, s'')$ in D_{4B}^i to either $z^*(s, s, s'')$ or $z_1(s, s'')$ in D_{7B}^i . By definition of $z_1(s, s'')$ and $z_2(s, s'')$, $z_1 R^i(s) z_2$, so $z^*(s, s, s'') = z_1(s, s'')$, and the outcome therefore does not change. If $\bar{m}^i = (x, s, s', -4, 0, s) \forall j \neq i$, $(\bar{m}^{-i}, m^i) \in D_{4B}^i$ and $(\bar{m}^{-i}, \bar{m}^i) \in D_{6B}^i$ so the outcome changes from $z_1(s, s')$ to $z_2(s, s')$. However, $z_1(s, s') R^i(s) z_2(s, s')$ so i is no better off. If $(\bar{m}^{-i}, m^i) \in D_{5B}^i$, then $(\bar{m}^{-i}, \bar{m}^i) \in D_{7B}^i$ and the outcome is either $z^*(s, s, s'')$ or $z_1(s, s'')$ in D_{7B}^i . Again, since $z^*(s, s, s'') = z_1(s, s'')$, the outcome does not change. Hence, \bar{m}^i does not weakly dominate m^i .

Case (iii): Any other deviation by i moves the joint message from D_1 to D_3^i . There are two cases: $x P^i(s) w^i(s)$ and $w^i(s) R^i(s) x$. If $x P^i(s) w^i(s)$, then i is strictly worse off by switching from m^i . If $w^i(s) R^i(s) x$, so that x is a worst element for i at s , consider $\bar{m}^j = (x, s, s, i, 0, s) \forall j \neq i$. Then,

$(\bar{m}^{-i}, m^i) \in D_{9A}^i$ while $(\bar{m}^{-i}, \bar{m}^i) \in D_{8B}^i$, and the outcome changes from $b^i(s)$ to $w^i(s)$. If $b^i(s)P^i(s)w^i(s)$, then i is strictly worse off by switching from m^i . If $w^i(s)R^i(s)b^i(s)$, then i is completely indifferent between all alternatives at s so m^i is not weakly dominated by any strategy. *Q.E.D.*

LEMMA 3: At any s , if m is a UNE and $m \notin D_1 \cup D_2$, then $g(m) \in F(s)$.

PROOF: Outside of $D_1 \cup D_2$, there are always at least $I - 1$ agents each of whom can unilaterally move the joint message to D_{10} and obtain their best element at s . Hence, if m is a Nash equilibrium and $m \notin D_1 \cup D_2$, it must be the case that $g(m)R^i(s)b^i(s)$ for at least $I - 1$ agents. No veto power then yields $g(m) \in F(s)$. Since $UNE(s) \subset NE(s)$, the conclusion follows. *Q.E.D.*

LEMMA 4: If $m \in D_1$ is a UNE at s' , then $g(m) \in F(s')$.

PROOF: Let $m^j = (x, s, s, 0, 0, s) \forall j$. If $s' = s$, then since $M_1 = \{(x, s) | x \in F(s)\}$, we trivially have $x \in F(s)$, so suppose $s \neq s'$ and $x \notin F(s')$. Then either (i) or (ii) of Definition 9 holds.

Suppose (i) holds. We claim that $\bar{m}^i = (x, s, s', -1, 2, s')$ weakly dominates m^i for agent $i_1(s, s')$, a contradiction. To see this, observe that the change from m^i to \bar{m}^i only affects the outcome when $(\bar{m}^{-i}, m^i) \in D_{4A}^i \cup D_{5A}^i \cup D_{8A}^i$. We analyze each case separately.

(1) If $(\bar{m}^{-i}, m^i) \in D_{4A}^i$, $(\bar{m}^{-i}, \bar{m}^i) \in D_{6A}^i \cup D_{7A}^i$. The outcome changes from $y_1(s, s'')$ in D_{4A}^i to $b^i(s')$ in D_{7A}^i (since $k^j = 0 \forall j \neq i$) or from $y_1(s, s')$ to $y_2(s, s')$ in D_{6A}^i . By definition, $b^i(s')R^i(s)y_1(s, s'')$ and $y_2R^i(s')y_1$.

(2) If $(\bar{m}^{-i}, m^i) \in D_{5A}^i$, $(\bar{m}^{-i}, \bar{m}^i) \in D_{7A}^i$, and the outcome changes from $w^i(s'')$ to $b^i(s')$ if $k' > \max_{j \neq i} k^j$ and remains at $w^i(s'')$ otherwise. When $s'' = s'$ and $k^j = 1 \forall j \neq i$, i is strictly better off since by definition of $i_1(s, s')$, $b^i(s')P^i(s')w^i(s')$.

(3) $(\bar{m}^{-i}, m^i) \in D_{8A}^i \Rightarrow (\bar{m}^{-i}, \bar{m}^i) \in D_{8A}^i$. The outcome either remains at $w^i(s'')$ or changes to $b^i(s')$. In either case, i is no worse off.

Suppose (ii) holds. We claim that $\bar{m}^i = (x, s, s', -3, 2, s')$ weakly dominates m for agent $i_2(s, s')$. To see this, observe that the change from m^i to \bar{m}^i only affects the outcome when $(\bar{m}^{-i}, m^i) \in D_{4B}^i \cup D_{5B}^i$. We analyze each case separately.

(1) If $(\bar{m}^{-i}, m^i) \in D_{4B}^i$, $(\bar{m}^{-i}, \bar{m}^i) \in D_{6B}^i \cup D_{7B}^i$. When $(\bar{m}^{-i}, m^i) \in D_{6B}^i$, then the outcome changes from $z_1(s, s')$ in D_{4B}^i to $z_2(s, s')$ in D_{6B}^i , and i is strictly better off since by definition, $z_2P^i(s')z_1$. If $(\bar{m}^{-i}, \bar{m}^i) \in D_{7B}^i$, the outcome either changes from $z_1(s, s'')$ to $z^*(s', s, s'')$ or remains $z_1(s, s')$. In either case, i is no worse off.

(2) If $(\bar{m}^{-i}, m^i) \in D_{5B}^i$, $(\bar{m}^{-i}, \bar{m}^i) \in D_{7B}^i$, and the outcome either changes from $z_1(s, s'')$ to $z^*(s', s, s'')$ or remains at $z_1(s, s'')$. In either case, i is no worse off. *Q.E.D.*

LEMMA 5: If $m \in D_2$ is a UNE at s' with $m_1^j = (x, s) \forall j$, then $x \in F(s')$.

PROOF: As in Lemma 4, we may suppose that $s' \neq s$ and $x \notin F(s')$. Then either (i) or (ii) of Definition 9 hold. In either case, we derive a contradiction.

If (i) holds, consider agent $i = i_1(s, s')$:

(a) If i is playing $m^i = (x, s, s, 0, 0, s)$, then the same argument as in Lemma 4 can be used to show that m^i is weakly dominated.

(b) If i is playing $m^i = (x, s, s'', -1, k', s')$, then $\bar{m}^i = (x, s, s'', -1, k' + 1, s')$ weakly dominates m^i . To see this, note that the change from m^i to \bar{m}^i only affects the outcome in D_{7A}^i and in D_{8A}^i . In D_{7A}^i , when $\bar{m}^j = (x, s, s', -2, k', s) \forall j \neq i$, the outcome changes from $w^i(s')$ to $b^i(s')$, and i is strictly better off. For any other \bar{m}^j , $j \neq i$, such that $(\bar{m}^{-i}, m^i) \in D_{7A}^i$, i is no worse off. If $(\bar{m}^{-i}, m^i) \in D_{8A}^i$, the outcome can change from $w^i(s'')$ to $b^i(s')$ or remain the same. In either case, i is no worse off. Hence $g(\bar{m}^{-i}, \bar{m}^i)R^i(s')g(\bar{m}^{-i}, m^i) \forall \bar{m}^{-i}$ and $g(\bar{m}^{-i}, \bar{m}^i)P^i(s')g(\bar{m}^{-i}, m^i)$ when $\bar{m}^i = (x, s, s', -2, k', s) \forall j \neq i$, so that \bar{m}^i weakly dominates m^i .

(c) If i is playing $m^i = (x, s, s'', -1, k', s')$ for $s'' \neq s'$, then changing from s^i to s' never hurts i , so \bar{m}^i given in (b) weakly dominates m^i .

(d) If i is playing $m^i = (x, s, s'', -3, k', s')$, we claim that if $z_2(s, \hat{s})P^i(s')z_1(s, \hat{s})$ for some \hat{s} such that $i = i_2(s, \hat{s})$, then $\bar{m}^i = (x, s, s'', -3, k' + 1, s')$ weakly dominates m^i . To see this, note that the change from m^i to \bar{m}^i only affects the outcome in D_{7B}^i . When $\bar{m}^j = (x, s, \hat{s}, -4, k', s) \forall j \neq i$, the outcome switches from $z_1(s, \hat{s})$ to $z^*(s', s, \hat{s}) = z_1(s, \hat{s})$ since $z_2(s, \hat{s})P^i(s')z_1(s, \hat{s})$, and i is strictly better off. Further, i is no worse off for any \bar{m}^{-i} such that $(\bar{m}^{-i}, m^i) \in D_{7B}^i$.

Suppose, then, that $z_1(s, \hat{s})R'(s')z_2(s, \hat{s})$ for all \hat{s} such that $i = i_2(s, \hat{s})$. In this case, $z^*(s', s, \hat{s}) = z_1(s, \hat{s})$ for all such \hat{s} , and we claim $\bar{m}' = (x, s, s', -1, 2, s')$ weakly dominates m' . To see this, observe that the outcome can only change in D_{4A}^i , D_{5A}^i , D_{6B}^i , D_{7B}^i , or D_{8A}^i . If $(\bar{m}^{-i}, m') \in D_{5A}^i$, then $(\bar{m}^{-i}, \bar{m}^{-i}) \in D_{7A}^i$. When $\bar{m}' = (x, s, s', -2, 1, s) \forall j \neq i$, the outcome changes from $w^i(s')$ to $b^i(s')$ and i is strictly better off. For any other m^{-i} such that $(m^{-i}, \bar{m}') \in D_{7A}^i$, i is no worse off. Elsewhere, it is easy to see that i is no worse off so \bar{m}' weakly dominates m' (see Palfrey and Srivastava (1986b) for details).

(e) If i is playing $m^i = (x, s, s'', -3, k^i, s^i)$, $s^i \neq s'$, then changing from s^i to s' never hurts i at s' , and the arguments in (d) apply.

To summarize (a)–(e), if (i) of Definition 9 holds, then m' is weakly dominated. Suppose then that (ii) of Definition 9 holds, and consider agent $i = i_2(s, s')$.

(f) If i is playing $m^i = (x, s, s'', -3, k^i, s^i)$, then $\bar{m}' = (x, s, s'', -3, k' + 2, s')$ weakly dominates m' . To see this, note that the outcome can only change in D_{7B}^i . Since $z_2(s, s')P^i(s')z_1(s, s')$, $z^*(s', s, s') = z_2(s, s')$, so when $\bar{m}' = (x, s, s', -4, k' + 1, s)$, the outcome changes from $z_1(s, s')$ to $z^*(s', s, s')$ and i is strictly better off. For any \bar{m}^{-i} such that $(\bar{m}^{-i}, \bar{m}') \in D_{7B}^i$, i is no worse off.

(g) If i is playing $m^i = (x, s, s'', -1, k^i, s^i)$, then $\bar{m}' = (x, s, s'', -1, k' + 2, s')$ weakly dominates m' . To see this, note that the outcome only changes in D_{8A}^i , i is strictly better off when $\bar{m}' = (x, s, s', -2, k' + 1, s) \forall j \neq i$, and i is no worse off anywhere else. Q.E.D.

Lemmas 1–2 show $F(s) \subset UNE(s)$, Lemmas 3–5 show $UNE(s) \subset F(s)$, so $F(s) = UNE(s) \forall s$. Finally, we indicate how the mechanism has to be modified when some or all agents have no or worst elements at some states. If agent i has no worst element, the outcomes in the various regions have to be changed as follows: in D_3^i and D_{6B}^i to something strictly worse for i at s than x ; in D_{5A}^i and D_{7A}^i to something strictly worse for i at s'' than both $y_1(s, s'')$ and $y_2(s, s'')$; in D_{8A}^i , $w^i(s'')$ should be replaced by x if $i \neq i_1(s, s'')$ and $z_2(s, s'')$ if $i = i_2(s, s'')$. Q.E.D.

REFERENCES

- ABREU, D., AND A. SEN (1990): "Subgame Perfect Implementation," *Journal of Economic Theory*, 50, 285–299.
- BLUME, L. (1986): "Lexicographic Refinements of Nash Equilibrium," Mimeo, University of Michigan.
- BRANDENBURGER, A., AND E. DEKEL (1986): "On an Axiomatic Approach to Refinements of Nash Equilibrium: A Characterization of Perfect Equilibrium," Mimeo, University of California—Berkeley.
- CRAWFORD, V. (1979): "A Procedure for Generating Pareto-Efficient Egalitarian-Equivalent Allocations," *Econometrica*, 47, 49–60.
- VAN DAMME, E. (1987): *Stability and Perfection of Nash Equilibria*. Berlin: Springer-Verlag.
- DASGUPTA, P., P. HAMMOND, AND E. MASKIN (1979): "The Implementation of Social Choice Rules: Some General Results on Incentive Compatibility," *Review of Economic Studies*, 46, 185–216.
- DEKEL, E., AND D. FUDENBERG (1987): "Rational Behavior with Payoff Uncertainty," Mimeo, MIT.
- GIBBARD, A. (1973): "Manipulation of Voting Schemes: A General Result," *Econometrica*, 41, 587–601.
- GROVES, T., AND J. LEDYARD (1988): "Incentive Compatibility Ten Years Later," in *Information, Incentives, and Economic Mechanisms*, T. Groves, R. Radner, and S. Reiter. Minneapolis, MN: University of Minnesota Press.
- JACKSON, M. (1991): "Bayesian Implementation," *Econometrica*, 59, 461–477.
- MASKIN, E. (1977): "Nash Equilibrium and Welfare Optimality," Mimeo, MIT.
- (1986): "The Theory of Implementation in Nash Equilibrium: A Survey," in *Social Goals and Social Organization: Volume in Memory of Elisha Pazner*. Cambridge: Cambridge University Press.
- McKELVEY, R. (1989): "Game Forms for Nash Implementation of General Social Choice Correspondences," *Social Choice and Welfare*, 6, 139–156.
- MOORE, J., AND R. REPULLO (1988): "Subgame Perfect Implementation," *Econometrica*, 56, 1191–1220.
- MOULIN, H. (1979): "Dominance Solvable Voting Schemes," *Econometrica*, 47, 1337–1352.
- PALFREY, T., AND S. SRIVASTAVA (1989a): "Implementation with Incomplete Information in Exchange Economies," *Econometrica*, 57, 115–134.
- (1987): "On Bayesian Implementable Allocations," *Review of Economic Studies*, 54, 193–208.

- (1986a): "Implementation in Exchange Economies Using Refinements of Nash Equilibrium," Mimeo, Carnegie-Mellon University.
- (1986b): "Nash Implementation Using Undominated Strategies," Social Sciences Working Paper #649, Caltech.
- (1989b): "Mechanism Design with Incomplete Information: A Solution to the Implementation Problem," *Journal of Political Economy*, 97, 668–691.
- POSTLEWAITE, A. (1986): "Implementation via Nash Equilibria in Economic Environments," in *Social Goals and Social Organization: A Volume in Memory of Elisha Pazner*, ed. by L. Hurwicz, D. Schmeidler, and H. Sonnenschein.
- POSTLEWAITE, A., AND D. SCHMEIDLER (1986): "Implementation in Differential Information Economies," *Journal of Economic Theory*, 39, 14–33.
- REICHELSTEIN, S. (1987): "A Note on Feasible Implementations," *Economic Letters*, 25, 315–318.
- REPULLO, R. (1987): "A Simple Proof of Maskin's Theorem on Nash Implementation," *Social Choice and Welfare*, 4, 39–41.
- SAIJO, T. (1988): "Strategy Space Reduction in Maskin's Theorem: Sufficient Conditions for Nash Implementation," *Econometrica*, 56, 693–700.
- SATTERTHWAITE, M. (1975): "Strategy-proofness and Arrow's Conditions: Existence and Correspondence Theorems for Voting Procedures and Social Welfare Functions," *Journal of Economic Theory*, 10, 187–217.
- SELTEN, R. (1975): "Reexamination of the Perfectness Concept for Equilibrium Points in Extensive Games," *International Journal of Game Theory*, 4, 25–55.
- WALKER, M. (1981): "A Simple Incentive Compatible Scheme for Attaining Lindahl Allocations," *Econometrica*, 49, 65–73.
- WILLIAMS, S. (1984): "Sufficient Conditions for Nash Implementation," IMA Preprint Series, University of Minnesota.